

REPORT DOCUMENTATION PAGE

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE (DD-MM-YYYY) 18-02-2009		2. REPORT TYPE Final Performance Report		3. DATES COVERED (From - To) August 1, 2008 - November 30, 2008	
4. TITLE AND SUBTITLE INDIVIDUAL DECISION-MAKING IN UNCERTAIN AND LARGE-SCALE MULTI-AGENT ENVIRONMENTS				5a. CONTRACT NUMBER FA9550-08-I-0429	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Prashant Doshi (PI), Adam Goodie (Co-PI)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The University of Georgia UGA Research Foundation, Inc 617-621 Boyd Graduate Studies Research Center Athens, GA 30602				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research (AFOSR) 875 N. Randolph Street, Room 3112 Arlington, VA 22203				10. SPONSOR/MONITOR'S ACRONYM(S) AFOSR/PKR I	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Research undertaken in this initial performance period developed: (a) the first set of generally applicable approximation methods for the finitely nested interactive POMDP (I-POMDP) framework, and (b) novel probabilistic graphical models called interactive dynamic influence diagrams (I-DIDs) that generalize the well-known DIDs to multiagent settings. These methods provide approximation techniques for decision making in complex multiagent settings in reduced time and space facilitating scalability. Experiments reveal that the approaches generate solutions of flexible quality proportional to the computational resources allocated. In the context of human decision making, this research showed that a strategic setting that was relatively simple, realistic and competitive increased the tendency in subjects to attribute higher levels of reasoning to others, which are consistent with typical levels of adversaries' reasoning. This complements previous experiments in less competitive settings which showed decision making to be reasonable only if other player is assumed to reason at a very low level. These data will be used to inform I-POMDP models and test their predictive behavior for descriptive validity. Outcomes of this research led to publications in several prestigious journals and conferences.					
15. SUBJECT TERMS decision making, multiagent settings, recursive reasoning, computational models, theory of mind					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 12	19a. NAME OF RESPONSIBLE PERSON Prof. Prashant Doshi
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code) 706-583-0827

FINAL PERFORMANCE REPORT

Contract/Grant title:

INDIVIDUAL DECISION-MAKING IN UNCERTAIN AND LARGE-SCALE MULTI-AGENT ENVIRONMENTS

Contract/Grant number:

FA9550-08-1-0429

Principal Investigator:

Prashant Doshi, University of Georgia

Reporting Period:

August 1, 2008 – November 30, 2008

Annual accomplishments (200 words max):

Research undertaken in this initial performance period developed: (a) the first set of generally applicable approximation methods for the finitely nested interactive POMDP (I-POMDP) framework, and (b) novel probabilistic graphical models called interactive dynamic influence diagrams (I-DIDs) that generalize the well-known DIDs to multiagent settings. These methods provide approximation techniques for decision making in complex multiagent settings in reduced time and space facilitating scalability. Experiments reveal that the approaches generate solutions of flexible quality proportional to the computational resources allocated. In the context of human decision making, this research showed that a strategic setting that was relatively simple, realistic and competitive increased the tendency in subjects to attribute higher levels of reasoning to others, which are consistent with typical levels of adversaries' reasoning. This complements previous experiments in less competitive settings which showed decision making to be reasonable only if other player is assumed to reason at a very low level. These data will be used to inform I-POMDP models and test their predictive behavior for descriptive validity. Outcomes of this research led to publications in several prestigious journals and conferences.

Archival publications during reporting period:

1. Adam Goodie, Prashant Doshi and Diana Young, "Two Level Recursive Reasoning by Humans Playing Sequential Fixed-Sum Games", *Twenty-first International Joint Conference on Artificial Intelligence (IJCAI)*, 6 pages, 2009, *submitted*
2. Yifeng Zeng and Prashant Doshi, "Speeding Up Exact Solutions of Interactive Dynamic Influence Diagrams Using Action Equivalence", *Twenty-first International Joint Conference on Artificial Intelligence (IJCAI)*, 6 pages, 2009, *submitted*
3. Prashant Doshi and Yifeng Zeng, "Improved Approximation of Interactive Dynamic Influence Diagrams Using Discriminative Model Updates", *Eight International Autonomous Agents and Multiagent Systems Conference (AAMAS)*, 8 pages, Budapest, Hungary, 2009, *to appear*

20090325303

4. Prashant Doshi, "Compact Approximations of Mixture Distributions for State Estimation in Multiagent Settings, short paper, *Eight International Autonomous Agents and Multiagent Systems Conference (AAMAS)*, 2 pages, Budapest, Hungary, 2009, to appear
5. Prashant Doshi and Piotr Gmytrasiewicz, "Monte Carlo Sampling Methods for Approximating Interactive POMDPs", *Journal of Artificial Intelligence Research (JAIR)*, 2009, 42 pages, to appear
6. Prashant Doshi, Yifeng Zeng and Qiongyu Chen, "Graphical Models for Interactive POMDPs: Representations and Solutions", *Journal of Autonomous Agents and Multiagent Systems (JAAMAS)*, Springer Publishing, Vol. 18(3):376-416, 2009.
7. Yifeng Zeng and Prashant Doshi, "An Information-Theoretic Approach to Model Identification in Interactive Influence Diagrams", *IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT)*, pp. 224-230, Sydney, Australia, 2008

Changes in research objectives, if any:

None

Change in AFOSR program manager:

None

Extensions granted or milestones slipped, if any:

None

Include any new discoveries, inventions, or patent disclosures during this reporting period:

None

A. Project Goals (abbreviated from the original proposal)

Goal 1. Identify ways to model the strategic reasoning of others in the interactive POMDP (I-POMDP) framework. We will investigate multiple open questions: (a) How do we decide the strategy level of an agent without being arbitrary? (b) What strategy levels should be ascribed to the other agent? (c) How do we consider the possibility that the other agent is more sophisticated (has deeper beliefs)?

Goal 2. Understand the sources of computational complexity within I-POMDPs and identify ways to mitigate their impact without significant losses in the optimality of the decision-maker. Motivated by the success of POMDP approximation schemes, we will identify the practical impediments to applying I-POMDPs to the large-scale settings that we have in mind. We will use this knowledge to develop a collection of new *approximation* techniques that will flexibly trade off optimality of the decisions for computational efficiency. These techniques will exploit notions of prescriptive similarity, decision-making heuristics, and the embedded structure in decision problems. The outcome of this research will be a collection of methods capable of generating decisions using bounded computing resources without significantly compromising the decision quality.

Goal 3. Incorporate the strategic behaviors and mental models of human agents using models that are psychologically plausible, empirically supported, and computationally tractable. This research will focus on modeling the behavioral data related to the *theory of mind* (TOM), in order to develop empirically informed computational models of strategic human decision making. TOM seeks to understand others' minds as possessing beliefs, desires and intentions similar to one's own mind. This research also proposes new empirical studies that enhance the understanding of strategic human behavior and contribute to TOM.

B. Activities and Results

Goal 1. Identify ways to model the strategic reasoning of others in the interactive POMDP (I-POMDP) framework.

Research on this goal has not yet begun and will be performed in the next period.

Goal 2. Understand the sources of computational complexity within I-POMDPs and identify ways to mitigate their impact without significant losses in the optimality of the decision-maker.

I. Monte Carlo sampling methods for *approximating* I-POMDPs

Interactive POMDPs (I-POMDPs) [Gmytrasiewicz and Doshi, 2005] are a generalization of POMDPs [Kaelbling et al., 1998] to multiagent settings and offer a principled framework for sequential decision making in uncertain multiagent settings. I-POMDPs are applicable to autonomous self-interested agents who locally compute what actions they should execute to optimize their preferences given what they believe while interacting with others with possibly conflicting objectives.

Continuing previous research, the PI developed the *first* set of generally applicable methods for computing approximately optimal policies for the finitely nested I-POMDP framework while demonstrating computational savings. Since an agent's belief is defined over other agents' models, which may be a complex infinite space, sampling methods which are able to approximate distributions over large spaces to arbitrary accuracy are a promising approach. Furthermore, sampling approaches allow us to *focus* resources on the regions of the state space that are considered more likely in an uncertain environment, providing a strong potential for computational savings. The research adopted the particle filter (PF) [Doucet et al., 2001] as a point of departure, and generalized the PF to the multiagent setting, resulting in the *interactive particle filter* (I-PF). The generalization is not trivial: Other agents are not simply treated as automata whose actions follow a fixed and known distribution. Rather, other agents are intentional – they possess beliefs, capabilities and preferences. Subsequently, the propagation step in the I-PF becomes more complicated than in the standard PF. In projecting the subject agent's belief over time, the other agent's belief must be projected, which involves predicting its action and anticipating its observations. Mirroring the hierarchical character of interactive beliefs, the interactive particle filtering involves sampling and propagation at each of the hierarchical levels of the beliefs. The research empirically demonstrated the ability of the I-PF to flexibly approximate the state estimation in I-POMDPs, and showed the computational savings obtained in comparison to a regular grid based implementation. However, as an identical number of particles are sampled at each nesting

level, the total number of particles and the associated complexity continues to grow exponentially with the nesting level. Figure 1 illustrates the operation of the I-PF.

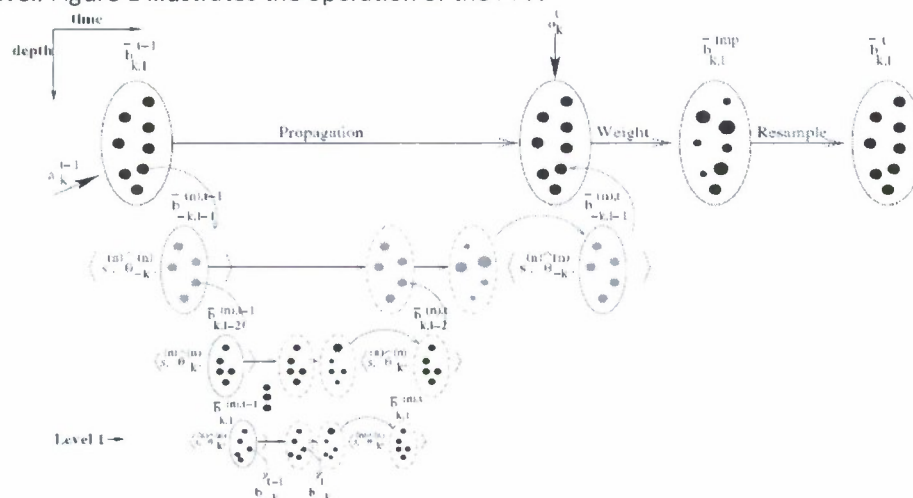


Figure 1: An illustration of the nesting in the I-PF. Colors black and gray distinguish filtering for the two agents. Because the propagation step involves updating the other agent's beliefs, particle filtering is performed on its beliefs. The filtering terminates when it reaches the level 1 nesting, where a level 0 belief update is performed for the other agent.

Empirical investigation of the performance of the I-PF showed that it approximates the exact state estimation closely. In addition to several toy problems, experiments were performed with a *UAV reconnaissance problem* in which the task of a UAV is to perform low-altitude reconnaissance of a potentially hostile theater that may be populated by other agents with conflicting objectives that serve as ground reconnaissance targets (see Fig. 2(a)).

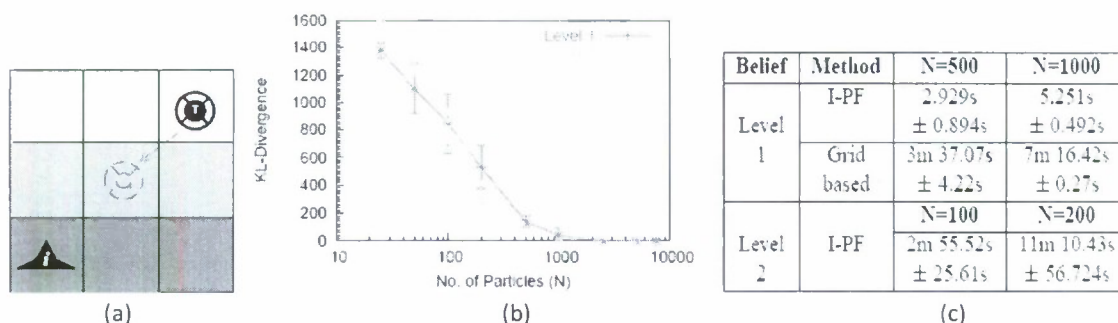


Figure 2: (a) The operating theater of the UAV i . The problem may be flexibly scaled by adding more targets and sectors. (b) The posterior obtained from the I-PF approaches the exact as the number of particles increase. (c) Comparison of the average running times of a numerical integration and I-PF implementations on the same platform (Xeon, 3.0GHz, 2GB RAM, Linux).

Combining the I-PF with value iteration on sample sets provides a general way to solve finitely nested I-POMDPs. The approximation method is *anytime* and is applicable to agents that start with a prior belief and optimize over finite horizons. Consequently, the method finds applications for online plan computation. Experimental analysis demonstrates (a) a reduction in error with increasing sample

complexity, and (b) savings in computation time when the approximation technique is used. Performance of the method on the UAV reconnaissance problem is shown in Fig. 3.

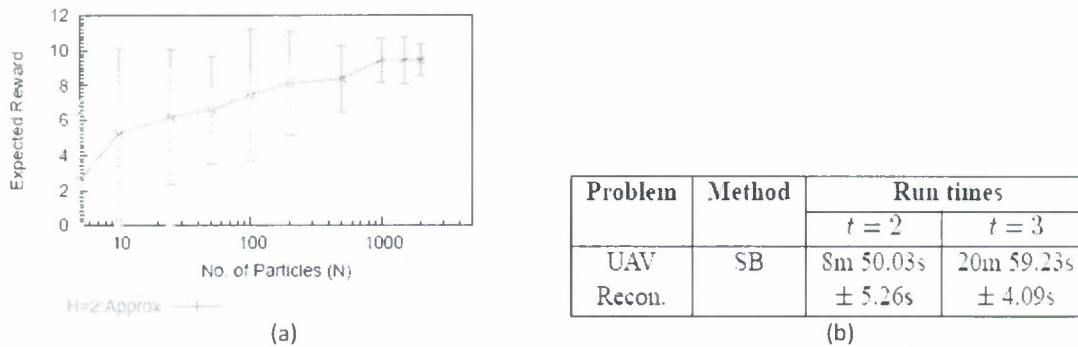


Figure 3: (a) Anytime performance profile for the UAV reconnaissance problem for horizons 2 and 3. Notice that the profile flattens when the number of particles reaches 1,000 and greater, thereby indicating that the corresponding average reward is close to optimal. (b) Run times for obtaining a policy on a Xeon 3.0 GHz, 2.0GB RAM and Linux. For horizon 2, we used 500 particles while 100 particles were used for horizon 3. As we see in (a), rewards of policies at these numbers of particles are not much less than the converged values.

- Personnel on this research: Prashant Doshi (PI), Xia Qu (grad. student)
- Publications during the grant period related to this research:
 1. Prashant Doshi, "Compact Approximations of Mixture Distributions for State Estimation in Multiagent Settings, short paper, *Eight International Autonomous Agents and Multiagent Systems Conference (AAMAS)*, 2 pages, Budapest, Hungary, 2009, to appear
 2. Prashant Doshi and Piotr Gmytrasiewicz, "Monte Carlo Sampling Methods for Approximating Interactive POMDPs", *Journal of Artificial Intelligence Research (JAIR)*, 42 pages, 2009, to appear

II. Interactive dynamic influence diagrams and its approximations

Enumerative representations of models often obscure important structure that is typically present in many realistic application settings. Graphical models such as *influence diagrams* (IDs) [Tatman and Shachter, 1990] offer a qualitative language that decomposes the state into *chance* (random) variables and dependencies between the variables. Algorithms for solving the models exploit the conditional independence between variables, and often consume less time and space in solving the problem compared to those that operate on traditional enumerative representations. Interactive dynamic influence diagrams (I-DIDs) may be viewed as graphical representations of I-POMDPs. They generalize DIDs (dynamic IDs), which are graphical counterparts of POMDPs to multiagent settings in the same way that I-POMDPs generalize POMDPs (Fig. 4).



Figure 4: The relationship between the four representations along two dimensions. The vertical dimension (dashed arrows) specifies the generalization from single agent to multiagent setting, while the horizontal dimension (solid arrows) is the mapping from enumerative to the graphical representation.

Analogous to DIDs, I-DIDs compactly represent the decision problem by mapping various variables into chance, decision and utility nodes, and denoting the dependencies between variables using directed arcs between the corresponding nodes. However, matters are more complex when we consider multiagent interactions that are extended over time, where predictions about others' future actions must be made using models that change as the agents act and observe. As shown in Fig. 5, I-DIDs address this gap by allowing the representation of other agents' models as the values of a special *model node*. In addition to the model node, I-DIDs differ from DIDs by having a dashed link, called the *policy link*, between the model node and a chance node, A_j , that represents the distribution over the other agent's actions given its model. In the absence of other agents, the model node and the chance node, A_j , vanish and I-DIDs collapse into traditional DIDs.

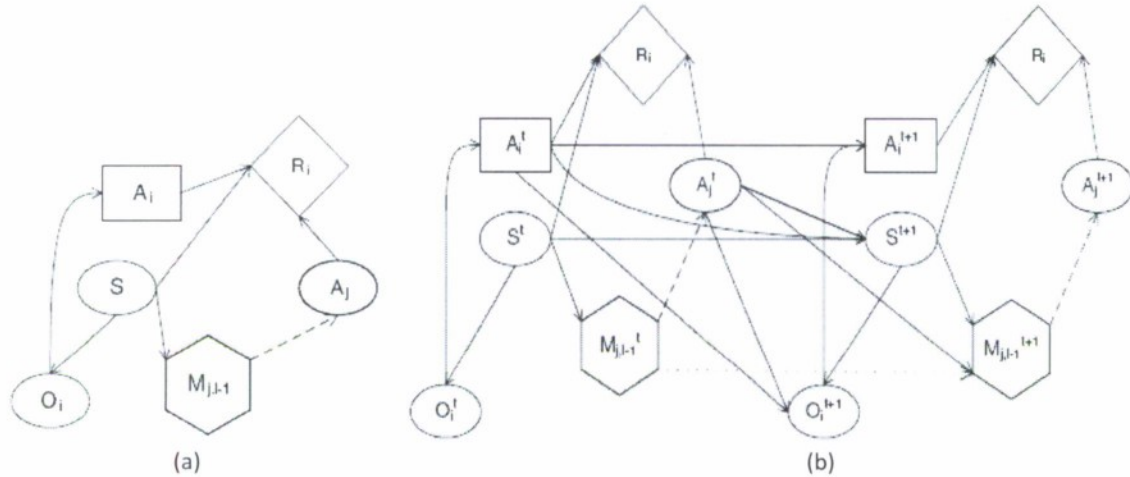


Figure 5: (a) A generic level 1 I-ID for agent i situated with one other agent j . The hexagon is the model node ($M_{j,i-1}$) and the dashed arrow is the policy link. Members of the model node could be I-IDs themselves or IDs ($m_{j,i-1}^1, m_{j,i-1}^2$; diagrams not shown here for simplicity) representing intentional models. (b) I-DID unrolled over two time horizons. The dotted arrow between the model nodes is the model update link.

Both other agents' models and the original agent's beliefs over these models are updated over time. Specifically, the update of the agent's belief over the models of others as the agents act and receive observations is denoted using a special link called the *model update link* that connects the model nodes between time steps. To facilitate understanding, we explicate the semantics of the model node and the model update link by showing how they can be implemented using the traditional dependency links between the chance nodes that constitute the model nodes (Fig. 6).

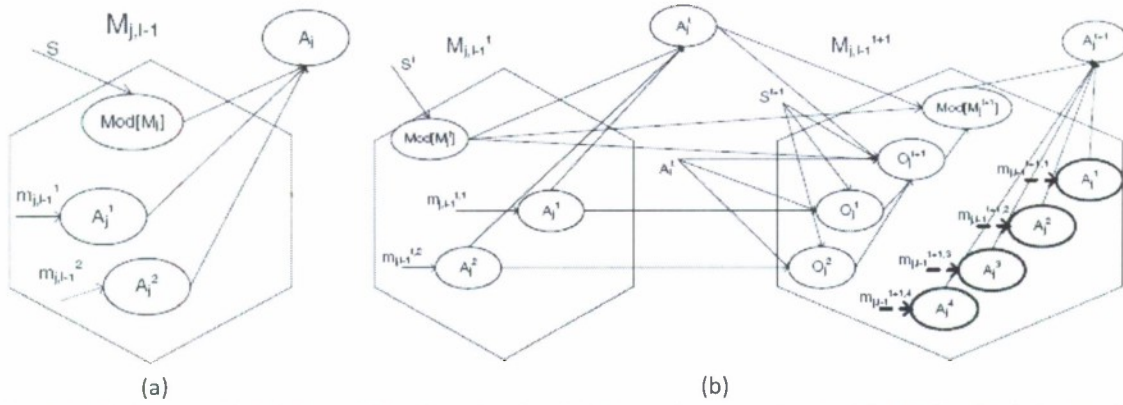


Figure 6: (a) Representing the model node and policy link using chance nodes and dependencies between them. The decision nodes of the lower-level I-DIDs or IDs ($m_{j,l-1}^1, m_{j,l-1}^2$) are mapped to the corresponding chance nodes (A_j^1, A_j^2), which is indicated by the dotted arrows. Depending on the value of the node, $\text{Mod}[M_j]$, the distribution of each of the chance nodes is assigned to the node A_j in its CPD. (b) Representing the model update link between model nodes using chance nodes and dependency links between them. Notice the growth in the number of models in the model node at $t + 1$ (highlighted in bold)

In order to illustrate the usefulness of I-DIDs, they were applied to three illustrative problems. In particular, they were used to demonstrate altruism and reciprocity in the public good game, the emergence of followership and leadership in the multiagent tiger problem, and mistrust in the online shopper's dilemma.

Expectedly, I-DIDs acutely suffer from both the curses of dimensionality and history. The PI developed multiple methods of reducing the dimensionality of the interactive state space and mitigate the impact of the curse of history that afflicts the modeled agents:

1. The first method, labeled as MC, limits and holds constant the number of models, $0 < K_{MC} < M$, where M is the possibly large number of candidate models of the other agents included in the model node. Using the insight that beliefs that are spatially close are likely to be behaviorally equivalent, the approach *clusters* the models of the other agents and selects representative models from each cluster. In this regard, the popular *k*-means clustering method is used, which gives an iterative way to generate the clusters. Intuitively, the clusters contain models that are likely to be behaviorally equivalent and hence may be replaced by a subset of representative models without a significant loss in the optimality of the decision maker. K_{MC} representative models are selected from the clusters and updated over time. Run times indicative of computational savings are shown in Fig. 7.

Problem	Method	Horizons			
		$t = 2$	$t = 3$	$t = 4$	$t = 5$
Multiagent	Exact	14.079s	33.142s	83.644s	*
Tiger	MC	4.532s	7.110s	10.512s	12.328s
Multiagent	Exact	14.234s	35.847s	99.236s	*
Machine maintenance	MC	8.500s	12.908s	18.688s	33.219s

Figure 7: Run times for exactly and approximately solving I-DID for different horizons. K_{MC} and M are equal to 50 and 100 respectively for both approximate and exact approaches (Pentium 4, 3.0GHz, 1GB RAM, WinXP). * Exact solutions ran out of memory.

2. The second method, labeled as DMU, significantly reduces the space of possible models of other agents that we need consider by discriminating between model updates. Specifically, at each time step, it selects only those models for updating which will result in predictive behaviors that are distinct from others in the updated model space. In other words, models that on update would result in predictions which are identical to those of existing models are not selected for updating. For these models, their revised probability masses are simply transferred to the existing behaviorally equivalent models. Intuitively, this approach improves on the previous one because it does not generate all possible models prior to selection at each time step; rather it results in minimal sets of models. In order to avoid updating all models, regions of the belief space are found so that models whose beliefs fall in these regions will be behaviorally equivalent on update. Note that these regions need not be in spatial proximity. Because obtaining the exact regions is computationally intensive, the method approximately obtains these regions by solving a subset of the models, $0 < K_{DMU} < M$, and utilizing their combined policies. Fig. 8 demonstrates the empirical performance of this approach on a toy problem and its significant improvement over the previous approach of model clustering.

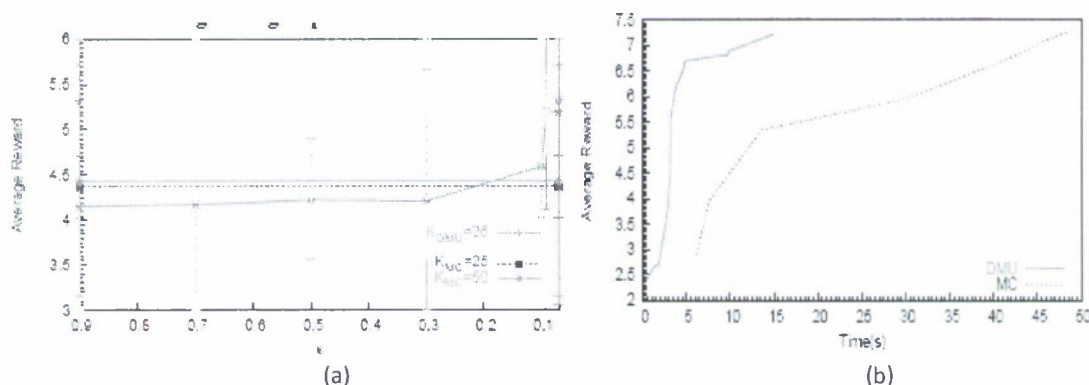


Figure 8: (a) Performance profiles for the multiagent tiger problem generated by executing policies obtained using discriminative model updates (DMU) on an I-DID of horizon 8; $M_0=100$. As K_{DMU} increases and epsilon reduces, the performance approaches that of the exact for given M_0 . We compare with model clustering (MC) for varying K_{MC} as well. Vertical bars represent the standard deviations. (b) Notice that an I-DID solved using DMU requires approximately an order of magnitude less time as the MC to produce comparable solutions.

- Personnel on this research: Prashant Doshi (PI)
- Publications during the grant period related to this research:
 1. Yifeng Zeng and Prashant Doshi, "Speeding Up Exact Solutions of Interactive Dynamic Influence Diagrams Using Action Equivalence", *Twenty-first International Joint Conference on Artificial Intelligence (IJCAI)*, 6 pages, 2009, *submitted*
 2. Prashant Doshi and Yifeng Zeng, "Improved Approximation of Interactive Dynamic Influence Diagrams Using Discriminative Model Updates", *Eight International Autonomous Agents and Multiagent Systems Conference (AAMAS)*, 8 pages, Budapest, Hungary, 2009, *to appear*
 3. Prashant Doshi, Yifeng Zeng and Qiongyu Chen, "Graphical Models for Interactive POMDPs: Representations and Solutions", *Journal of Autonomous Agents and Multiagent Systems (JAAMAS)*, Springer Publishing, Vol. 18(3):376-416, 2009.
 4. Yifeng Zeng and Prashant Doshi, "An Information-Theoretic Approach to Model Identification in Interactive Influence Diagrams", *IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT)*, pp. 224-230, Sydney, Australia, 2008

Goal 3. Incorporate the strategic behaviors and mental models of human agents using models that are psychologically plausible, empirically supported, and computationally tractable.

I. Two level recursive reasoning by humans playing strategic, sequential games

Strategic recursive reasoning of the form *what do I think that you think that I think* (and so on) arises naturally in multiagent settings. For example, a robotic uninhabited aerial vehicle (UAV)'s decision may differ if it believes that its reconnaissance target believes that it is not being spied upon in comparison to when the UAV believes that its target believes that it is under surveillance. Evidence of recursive reasoning in humans and investigations into the level of such reasoning is relevant to multiagent decision making in *mixed* settings. In particular, these results are directly applicable to computational frameworks such as the recursive modeling method, interactive POMDP [Gmytrasiewicz and Doshi, 2005] and cognitive ones such as theory of mind (TOM) [Dunbar, 1998] that ascribe intentional models of behavior to other agents.

Initial investigations into ascertaining the depth of strategic reasoning of humans by Stahl and Wilson [1995], and more recently, by Hedden and Zhang [2002] and Ficci and Pfeffer [2008] show that humans generally operate at only first or second level of recursive reasoning. Typically the first level, which attributes no recursive reasoning to others, is more prominent. Evidence of these shallow levels of reasoning is not surprising, as humans are limited by bounded rationality.

This research hypothesized that a strategic setting that is realistic, competitive and includes tangible incentives would increase participants' tendency to attribute levels of reasoning to others that reflect individuals' actual level of reasoning. A large study was conducted with human subjects to test the hypothesis. The study utilized a task that resembled the two-player sequential game as used by Hedden and Zhang but made simultaneously simpler and more competitive by incorporating fixed-sum outcomes and monetary incentives (see Fig. 9). Subjects played the game against a computer opponent, although they were led to believe that the opponent was human. Different groups of subjects were paired against an opponent that used no recursive reasoning (zero level) and opposite one that used first-level reasoning. The realism of the task was also manipulated between participants, with one group experiencing the task described abstractly while the other experienced a task that was structurally identical but described using a realistic cover story involving UAV reconnaissance.

Methodology In order to test different levels of recursive reasoning, the computer opponent (player II) was designed to play a game in two ways: (i) If player I chooses to move, II decides on its action by simply choosing between the outcomes at states B and C in Fig. 9(b) rationally. Therefore, II is a zero-level player and is called *myopic*. (ii) If player I chooses to move, the opponent decides on its action by reasoning what player I will do rationally. Based on the action of I, player II will select an action that maximizes its outcomes. Thus, player II is a first-level player, and is called *predictive*.

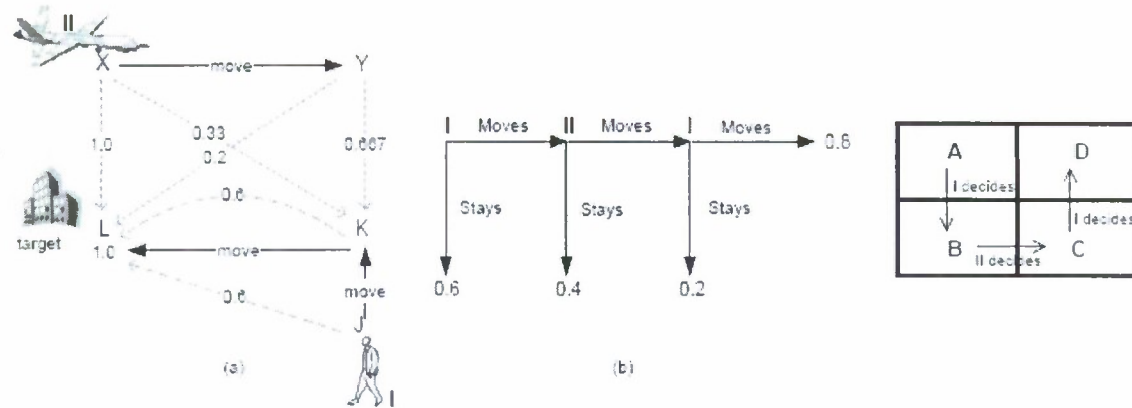


Figure 9: (a) A spatial visualization of the game where player *I* is a human intending to gather information about a target. Player *II* is a human-controlled UAV aiming to hinder *I* from gathering the critical information. The dashed arrows and probabilities indicate the chances of *I* gathering information or *II* hindering its access. (b) Centipede representation of our game with the outcomes as the probabilities of success of player *I*. It is a fixed-sum game and the remaining probability is the chance of success of player *II* (failure of player *I*). States of the game are indicated by the letters, A, B, C and D. Arrows denote the progression of play in the game. An action of *move* by each player causes a transition of the state of the game.

Notice that the rational choice of players in the game depends on the preferential ordering of states of the game rather than actual values. Let $a < b$ indicate that the player whose turn it is to play prefers state b over a , and because the game is purely competitive, the other player prefers state a over b . Games that exhibit a preference ordering of $D < C < B < A$ and $A < B < C < D$ for player *I* are trivial because player *I* will always opt to stay in the former case and move in the latter case, regardless of how *II* plays. Furthermore, consider the ordering $C < A < B < D$ for player *I*. A myopic opponent will choose to move while a predictive opponent will stay. However, in both these cases player *I* will choose to move. Thus, games whose states display a preferential ordering of the type mentioned previously are *not diagnostic* – regardless of whether player *I* thinks that opponent is myopic or is predictive, *I* will select the same action precluding a diagnosis of *I*'s level of recursive reasoning. Of all the 24 distinct preferential orderings among states that are possible, only one is diagnostic: $C < B < A < D$. For this ordering, player *I* will move if it thinks that the opponent is myopic, otherwise *I* will stay if the opponent is thought to be predictive. Note that the game in Fig. 9 follows this preference ordering.

Batches of participants played the game on computer terminals with each batch having an even number of players. Each batch was divided into two groups and members of the two groups were sent to different rooms. This was done to create the illusion that each subject was playing against a member of the other group, although the opponent was in reality a computer program. This deception was revealed to the subjects during debriefing. Each subject experienced an initial *training phase* of at least 15 games that were trivial or not diagnostic. These games served to acquaint the participants with the rules and goal of the task without unduly biasing them about the behavior of the opponent. Participants who failed to choose the rational actions in any of the previous 5 games after the 15-game training phase continued with new training games until they met the criterion of no rationality errors in the 5 most recent games. Those who failed to meet this criterion after 25 total training games did not advance to the test phase, and were removed from the study. In the *test phase*, each subject experienced 40 games instantiated with outcome probabilities that exhibited the diagnostic preferential ordering of $C < B < A < D$ for player *I*. The 40 critical games were divided into 4 blocks of 10 games each. In order to

avoid subjects developing a mental set, we interspersed these games with 40 that exhibited the orderings, $C < A < B < D$ and $D < B < A < C$. The latter games not only serve to distract the participants but also function as “catch” trials allowing us to identify participants who may not be attending to the games. They also led to better balancing of stay and switch trials. Approximately half the participants played against myopic opponents while the remaining played against predictive ones. In each category, approximately half of the participants were presented with just the Centipede representation of the games with probabilistic payoffs and no cover story, which is labeled as the *abstract* version. Remaining participants in the category were presented with the UAV cover story and the associated picture in Fig. 9(a), including the Centipede representation. This is labeled as the *realistic* version. About half of all participants also experienced a screen asking them what they thought the opponent would play and their confidence in the prediction, for some of the games. Participants received a monetary incentive of 50 cents for every correct action that they chose in a game. This resulted in an average payout of approximately \$30 per participant.

Results The study spanned a period of *three months* from September through November 2008. As mentioned before, each of 162 human subjects initially played a series of 15 games in order to get acquainted with the fixed-sum and complete information structure, and objectives of the task at hand. After this initial phase, participants who continued to exhibit errors in any of the games up to 25 total games were eliminated. 26 participants did not progress further in the study. These participants either failed to understand how the game is played or exhibited excessive irrational behavior, which would have affected the validity of the results of this study.

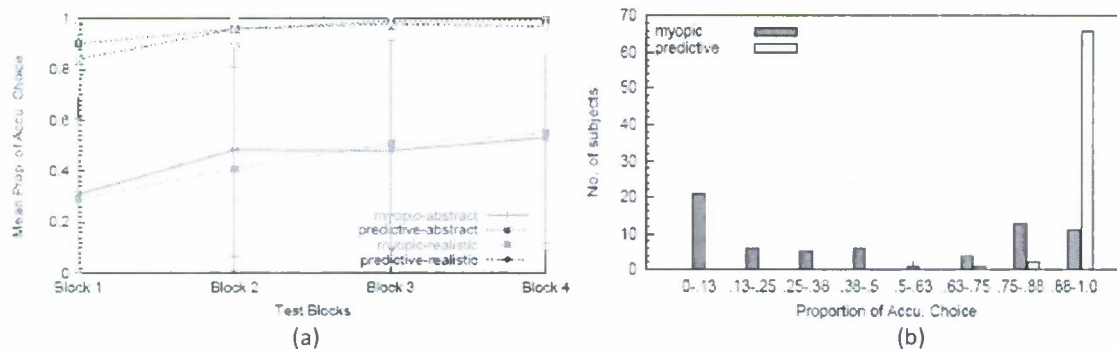


Figure 10: (a) Mean proportion of accurate choices of the participants for all conditions across test blocks. Notice that subjects generally expected their opponents to play at first level far more than at zero level. (b) Count of participants who played myopic or predictive opponents and grouped according to different proportions of accurate choice.

For each participant, we measured the proportion of times that the subject played accurately in each test block. We define an *accurate choice* as the action choice which is rational given the type of opponent. In Fig. 10(a), we show the mean proportion of accurate choices across all participants in each of the 4 groups. Two group-level findings are evident from the results in Fig. 10(a): First, the mean proportion of accurate choices is significantly higher when the opponent is predictive as compared to when it is myopic. Student t-tests with p-values < 0.0001 confirm that participants playing against predictive opponents have statistically significant higher proportions of accuracy compared to myopic opponents across all test blocks. The higher proportions of accurate choice when the opponent was predictive, in conjunction with lower proportions when the opponent was myopic, suggest that subjects predominantly displayed second-level reasoning when responding. They expected the opponent to

reason about their subsequent play (first-level reasoning) and acted accordingly. Second-order reasoning appeared to be the clear majority default strategy, and those for whom first-order reasoning was accurate showed only partial learning over 40 critical trials. Second, no significant difference in the mean proportions between abstract and realistic versions of the tasks is evident across any test block from Fig. 10(a). This is regardless of whether the opponent is myopic or predictive. Student t-tests with very low p-values revealed no statistical significance in the difference between the proportions, either overall or in any test block. The lack of any significant difference in the accuracy of the choices seems to suggest that our cover story neither confounded the participants nor clarified it further in an intuitive sense. We speculate that the indifference is due to, (i) the Centipede representation though abstract being sufficiently clear to facilitate understanding of this simple game, and thus (ii) subjects playing the games with high accuracy leaving little room for improvement, at least in the predictive groups. Finally, in Fig. 10(b), we detail the number of participants whose actions across all games fell into different bins of proportion of accurate choice. Fig. 10(b) reveals that about 83% of the 67 participants who played against a myopic opponent had proportions less than 0.875. In comparison, only about 4% of the 69 participants who played a predictive opponent exhibited such proportions of accurate choice.

Conclusions Data collected on the decisions of the participants indicate that:

1. Subjects responded accurately significantly more often when the opponent displayed first-level reasoning than when the opponent was at zero level. Learning with a first-order opponent was much faster and more complete than learning with a zero-order opponent.
2. No significant difference in the accuracy of the decisions was noticed between abstract and realistic task settings.

Thus, this study reveals clear evidence that higher levels of recursive reasoning could be observed in humans under simpler and more competitive settings with tangible incentives. However, there is room for future research on whether increased realism contributes to deeper levels of strategic reasoning.

- Personnel on this research: Adam Goodie (Co-PI), Prashant Doshi (PI), Diana Young (grad. student)
- Publications during the grant period related to this research:
 1. Adam Goodie, Prashant Doshi and Diana Young, "Two Level Recursive Reasoning by Humans Playing Sequential Fixed-Sum Games", *Twenty-first International Joint Conference on Artificial Intelligence (IJCAI)*, 2009, 6 pages, submitted

C. Planned research in the next period

- I. Modeling of strategic behavioral data using I-POMDPs

Two sets of data on human decision making in strategic settings are available. First set is available from [Hedden and Zhang, 2002], and includes the decisions of 70 participants playing strategic *general-sum* games. These data show that subjects initially tend to assign a zero-level of reasoning to opponents but gradually learn to assign a higher level if the opponent is reasoning at that level. The second set of data is available from the study conducted in this project. It includes the decisions of 162 participants playing strategic *fixed-sum* games with monetary incentives. This data shows that subjects predominantly attributed a higher-order of recursive reasoning to others by default, and showed faster and more complete learning in response to higher-order reasoning opponents than in response to lower-order reasoning opponents.

Because the payoff structure of the games is different in the two studies, these two sets of data complement each other. A computational model of this decision making should include the ability to model others at different levels of reasoning with differing priors and be able to learn the opponent model over time. This information is then translated into a decision. I-POMDPs allow all of these aspects. Importantly, the computational model should be sufficiently general to ascribe low levels of reasoning to opponents in general-sum games and switch to higher levels if the game is simpler and more competitive. Data from the studies will be used to inform the learning rate, the rationality of the decisions and levels of recursive reasoning in I-POMDPs.

II. Studies for validation and exploration of the role of partial observability

Two short studies will be conducted in order to replicate the results of the previous study and those reported by Hedden and Zhang [2002], under conditions where no monetary incentives are provided to the participants. These studies will serve to reinforce the validity of the results obtained previously under different conditions, within the context of the Co-PI and PI's laboratory settings.

A new study will be held in which the scenario will be made more complex, to include the possibility that the UAV's motions cannot be perfectly observed by the participants. This is of practical military importance, as the concealment of an enemy's movements is typically a priority for enemy planners. Data from this study will help in determining whether higher orders of recursive reasoning will continue to be observed under conditions of partial observability and whether partial observability affects the rate at which opponents' models are learnt and the completeness of the learning.

D. References

- [Doucet et al., 2001] Arnaud Doucet, Nando D. Freitas and Neil Gordon (eds.). *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2001.
- [Dunbar, 1998] Robin Dunbar. Theory of mind and evolution of language. In *Approaches to the Evolution of Language*. Cambridge University Press, 1998.
- [Ficici and Pfeffer, 2008] Sevan Ficici and Avi Pfeffer. Modeling how humans reason about others with partial information. In *Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 315–322, 2008.
- [Gmytrasiewicz and Doshi, 2005] Piotr Gmytrasiewicz and Prashant Doshi. A framework for sequential planning in multiagent settings. *Journal of Artificial Intelligence Research*, 24, 49–79, 2005.
- [Hedden and Zhang, 2002] Trey Hedden and Jun Zhang. What do you think i think you think?: Strategic reasoning in matrix games. *Cognition*, 85:1–36, 2002.
- [Kaelbling et al., 1998] Leslie Kaelbling, Michael Littman and Anthony Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134, 1998.
- [Stahl and Wilson, 1995] Dale Stahl and Paul Wilson. On player's models of other players: Theory and experimental evidence. *Gomes and Econ. Behavior*, 10:218–254, 1995.
- [Tatman and Shachter, 1990] J. A. Tatman and Ross D. Shachter. Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(2), 365–379, 1990.